

v0项目报告

项目背景与内容

本项目是《Engenharia do Conhecimento》课程的一个小型工程项目，主要任务是：

- 基于给定的 COVID-19 数据集，完成探索性数据分析（EDA）；
- 构建针对不同目标（O1-O4）的预测模型；
- 进行特征筛选、模型训练、优化与总结。

本项目并非完全自主设计，而是参考了课堂讲义与项目说明要求进行开发。

同时，我额外撰写了一份**个人版介绍文档**，在notebook/report部分，用于更好地解释代码逻辑。

开发过程与阶段小结

• EDA探索阶段

初期主要围绕数据的基本情况进行快速试探和筛选。理解到**探索性数据分析是一种博弈过程**：不断探索、快速假设与验证，确定有效特征，并识别数据中潜在的问题。

• OBJ建模阶段

初次处理OBJ相关任务时感到一定的困惑，但坚持先动手实践，再逐步加深理解。过程中回顾了TP课上学过的知识，虽然一开始生疏，但逐渐找回了感觉。

• 开发策略调整

一度尝试进行模块化开发，但在Notebook环境下，模块化效果反而适得其反。最终决定：**直接冲刺完整版**，以连贯的流程完成整个开发任务。

• 取舍与优化意识

在深入理解数据后，意识到某些特征虽存在，但对结果贡献极小。决定有选择地**舍弃干扰特征**，以提升整体分析与建模效率。

• 心态波动与调整

在中途开发过程中，面对繁琐细节与高强度要求，出现了疲惫和烦躁感。通过频繁审视讲义、审视代码和调整心态，逐步坚持完成。

当前完成度

- 已完成 EDA 数据探索与清洗；
 - 建立了针对 O1（死亡预测）、O2（年龄预测）、O3（死亡患者年龄预测）的初步模型；
 - 完成了初步的特征分析与总结；
 - 生成了多语言版本的报告材料（中文、英文、葡语）；
 - 保存了主要模型文件，以便后续调用与测试。
-

已知问题与后续计划

- 缺失值填充部分较为粗糙，主要使用了众数填充，未采用更高级的插补方法（如 KNN）；
 - 代码结构存在不一致，部分片段较为冗长，仍需整理优化；
 - 由于时间与精力限制，部分EDA可视化与深入统计分析未能完全展开；
 - 后续希望在空闲时间进一步回归，完善EDA深度与模型优化。
-

小结与感想

虽然本次v0版本尚不完美，但整体完成了任务要求，且在过程中积累了宝贵经验。

尤其是面对疲劳和压力时，能坚持完成，已是一种成长。

这次开发也让我更加理解了数据分析与机器学习项目的完整流程。未来希望能在不依赖过多辅助工具的情况下，独立完成更具深度的作品。

未来继续加油。

✅ 当前版本：v0

✅ 报告撰写日期：2025年5月4日